Journal of Chromatography A, 1270 (2012) 269–282

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Preliminary effects of real-world factors on the recovery and exploitation of forensic impurity profiles of a nerve-agent simulant from office media☆

Carlos G. Fraga[a,*], Landon H. Sego[a], Jamin C. Hoggard[b], Gabriel A. Pérez Acosta[a], Emilie A. Viglino[b], Jon H. Wahl[a], Robert E. Synovec[b]

[a] Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352, USA
[b] Department of Chemistry, Box 351700, University of Washington, Seattle, WA 98195, USA

ARTICLE INFO

Article history:
Received 13 August 2012
Received in revised form 25 October 2012
Accepted 25 October 2012
Available online 2 November 2012

Keywords:
Chemical forensics
Impurity profiling
Chemical threat agents
DMMP

ABSTRACT

Dimethyl methylphosphonate (DMMP) was used as a chemical threat agent (CTA) simulant for a first look at the effects of real-world factors on the recovery and exploitation of a CTA's impurity profile for source matching. Four stocks of DMMP having different impurity profiles were disseminated as aerosols onto cotton, painted wall board, and nylon coupons according to a thorough experimental design. The DMMP-exposed coupons were then solvent extracted and analyzed for DMMP impurities by comprehensive 2D gas chromatography/mass spectrometry (GC × GC/MS). The similarities between the coupon DMMP impurity profiles and the known (reference) DMMP profiles were measured by dot products of the coupon profiles and known profiles and by score values obtained from principal component analysis. One stock, with a high impurity-profile selectivity value of 0.9 out of 1, had 100% of its respective coupons correctly classified and no false positives from other coupons. Coupons from the other three stocks with low selectivity values (0.0073, 0.012, and 0.018) could not be sufficiently distinguished from one another for reliable matching to their respective stocks. The results from this work support that: (1) extraction solvents, if not appropriately selected, can have some of the same impurities present in a CTA reducing a CTA's useable impurity profile, (2) low selectivity among a CTA's known impurity profiles will likely make definitive source matching impossible in some real-world conditions, (3) no detrimental chemical–matrix interference was encountered during the analysis of actual office media, (4) a short elapsed time between release and sample storage is advantageous for the recovery of the impurity profile because it minimizes volatilization of forensic impurities, and (5) forensic impurity profiles weighted toward higher volatility impurities are more likely to be altered by volatilization following CTA exposure.

1. Introduction

To date we have found that different commercial stocks of synthetic chemicals such as dimethyl methylphosphonate (DMMP) have impurity profiles that are stock specific [1–3]. An impurity profile is a vector of data with each data point representing the signal intensity or concentration of an impurity obtained by chemical analysis of a sample. Typically, as in this paper, the signal intensities are normalized by dividing each impurity's signal intensity (e.g., chromatographic peak area) by the sum of all impurity intensities in the original impurity vector so that the sum of all intensities equals one. Impurity profiles have been demonstrated as a potential tool for associating different chemical samples according to source, i.e., chemical samples from the same source (e.g., stock, synthesis route, or geographic region) have statistically indistinguishable impurity profiles that are distinguishable from other sources to an extent [1–5]. This is potentially useful in forensic investigations involving chemical threat agents (CTAs) used in crimes and terrorist attacks by linking CTA samples from a crime scene to a CTA or starting materials obtained from the actual perpetrators. Recently, impurity profiles obtained from various batches of the chemical warfare agent sarin have demonstrated the ability to trace sarin to its specific starting material [1]. While the potential forensic value of impurity profiles has been investigated, their usefulness in cases when a CTA has been exposed to different matrices under real-world conditions has not been studied. Herein, we take a first look at the effects of some real-world factors on impurity profiles to get a better assessment of profile robustness and to understand the limits and criteria for source matching using impurity profiles.

One probable real-world scenario is an aerosol release of a CTA in an office building. In such an event, the aerosolized CTA deposits

☆ Presented at the 36th International Symposium on Capillary Chromatography and Electrophoresis and 9th GC × GC Symposium, Riva del Garda, Italy, 27 May–1 June 2012.

* Corresponding author. Tel.: +1 509 371 7589; fax: +1 509 375 2227.
E-mail address: carlos.fraga@pnnl.gov (C.G. Fraga).

0021-9673/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chroma.2012.10.053

itself onto office media that are later sampled and analyzed by forensic investigators. If a CTA's impurity profile is going to be useful in a forensic investigation, it needs to remain substantially unaltered during the time the CTA is disseminated, deposited, sampled, stored, and analyzed. Herein, we selected DMMP as a CTA simulant because it is an accepted surrogate and key precursor for G agents such as sarin [6], and its impurity profile has been investigated but not tested against real-world factors [3]. For the first time, this paper investigates the influence of eight real-world factors on the forensic utility of a CTA's impurity profile: (I) profile pattern, (II) aerosol-release duration, (III) plume duration, (IV) exposed matrix, (V) elapsed time from CTA deposition to sample storage, (VI) sample storage temperature, (VII) sample storage volume, and (VIII) extraction solvent. In essence, these factors are likely to be encountered when a CTA is released as an aerosol and the exposed surrounding media are then sampled, stored, and solvent extracted for chromatographic-mass spectrometric analysis. CTA decontamination was not addressed in this study, but was addressed in another study and not shown to have a noticeable influence on a CTA's impurity profile under the reported conditions [1]. Of course there are other factors, but those used here provide an initial assessment of impurity profiling for forensic applications under real-world conditions.

In this paper, the above eight factors were investigated in three studies (study 1, 2, and 3) that together tested the effects of these factors on the recovery and source matching of impurity profiles from different DMMP stocks disseminated onto typical office media. Comprehensive 2D gas chromatography/mass spectrometry (GC × GC/MS) was used to provide impurity profiles while the recovery of an impurity profile was measured by assessing the similarity between the profile from a DMMP-exposed coupon and its known DMMP stock profile. For the first time, the dot product was used to measure similarity among impurity profiles which proved useful in revealing and quantifying factor effects. Source matching was measured as the fraction of exposed coupons that were correctly matched to their DMMP stocks using all the tested DMMP stocks as source options. Finally, for the first time, the use of the analytical figure of merit known as selectivity was demonstrated as a potential way to measure the quality or forensic value of an impurity profile.

2. Theory

2.1. Profile similarity

Dot-product based algorithms are proven and typically used for measuring the similarities among mass spectra for chemical identification [7]. In this paper, we measured similarity using the normalized dot product of a coupon's impurity profile and reference DMMP stock profile. The normalized dot product is:

$$d = \left(\frac{\mathbf{r}}{||\mathbf{r}||}\right)^T \left(\frac{\mathbf{v}}{||\mathbf{v}||}\right) \quad (1)$$

where superscript T represents transpose, $||\ ||$ represents the Euclidean norm, $\mathbf{r}$ ($J \times 1$) is a reference impurity profile consisting of J impurities, $\mathbf{v}$ is like $\mathbf{r}$, but is the observed coupon profile to be matched to $\mathbf{r}$. The result, d, is a scalar value between 0 and 1, where a perfect match between impurity profiles yields $d = 1$, but two completely different (orthogonal) profiles yields $d = 0$.

Profile similarities can also be quantified and visualized through a scores plot obtained from principal component analysis (PCA) of impurity profiles. PCA is widely accepted and has been successfully used to illustrate similarities among CTA impurity profiles [5]. PCA converts a data matrix of impurity profiles into a scores matrix and loadings matrix for the selected principal components [8]. Impurity profiles that cluster closer together in scores space are more similar than those that are farther apart on the scores plot.

2.2. Profile selectivity

The uniqueness of an impurity profile relative to others can be measured by calculating its selectivity value. Selectivity for data vectors (e.g., impurity profiles) and higher-order data is defined by Faber et al. [9] and is a scalar that ranges from 0 to 1. In this paper, 1 indicates no overlap between the impurity profile of interest and all other profiles; 0 means complete overlap (i.e., no uniqueness) between the impurity profile of interest and all others. The key advantage of calculating selectivity is that it objectively reduces the complexity of comparing several impurity profiles by providing a single score for each profile. The calculation of selectivity first involves determining the net analyte signal for the impurity profile of interest. The net analyte signal is that portion of the impurity profile that is unique from all other profiles. It is obtained by the product of an orthogonal projection matrix and the impurity profile of interest:

$$\mathbf{r}^* = (\mathbf{I} - \mathbf{S}\mathbf{S}^+)\mathbf{r} \quad (2)$$

where $\mathbf{r}^*$ ($J \times 1$) is the net analyte signal of the impurity profile of interest $\mathbf{r}$, $\mathbf{I}$ ($J \times J$) is the identity matrix, $\mathbf{S}$ ($J \times [K\text{-}1]$) is the matrix consisting of all profiles minus the profile of interest, K is the total number of impurity profiles, and $\mathbf{S}^+$ is the pseudo-inverse of $\mathbf{S}$. If an impurity profile $\mathbf{r}$ is completely unique from all other profiles in $\mathbf{S}$, then its net analyte signal $\mathbf{r}^*$ will equal $\mathbf{r}$. If r is only partially unique from all other profiles, then $\mathbf{r}^*$ will not equal $\mathbf{r}$ and the two profiles will look different when viewed as bar graphs because $\mathbf{r}^*$ is missing portions from $\mathbf{r}$. Selectivity quantifies any differences between $\mathbf{r}^*$ and $\mathbf{r}$ through the ratio of the Euclidean norms ($||\ ||$) for $\mathbf{r}^*$ and $\mathbf{r}$:

$$\text{selectivity} = \frac{||\mathbf{r}^*||}{||\mathbf{r}||} \quad (3)$$

where a selectivity value of 1 means $\mathbf{r}^*$ equals $\mathbf{r}$ or the impurity profile of interest is totally unique; 0 means $\mathbf{r}^*$ is a vector of zeros or the impurity profile of interest has no uniqueness.

2.3. Experimental design and statistical analysis

The use of a screening study prior to a more in-depth study is a good practice for researchers trying to determine the influence of several factors on a phenomenon of interest. Herein, a $2^{(7-4)}_{\text{III}}$ design [10,11] (or two-to-the-seven minus four, resolution three fractional factorial design) was used to measure the main effects of seven factors that could potentially affect an impurity profile. The main effect for a factor is the difference in the measured variable (e.g., normalized dot product) when changing a factor's value between two levels, averaged across the levels of all other factors. The experimental design permits the testing of a maximum number of factors at two levels using only eight experiments. Its major limitation is that main effects are confounded with (or indistinguishable from) two-factor interactions. In general, it is prudent to include a "dummy" factor which has no expected influence on the measured outcome variable. The dummy factor is helpful for identifying a real factor that does not have a measurable effect because its main effect is less than that of the dummy factor [12]. ANOVA (analysis of variance) can then be used to provide insight as to what factors have a significant effect by using the pooled variability of those factors with main effects equal to or less than the dummy factor as experimental error [10]. The p-values and F-ratios from the ANOVA are then used to justify what factors are significant and worthy to be studied in depth.

In terms of an in-depth study, a split-plot design is often used when some factors in the experiment are more difficult to change